

PIXEL-RESOLUTION DTM GENERATION FOR THE LUNAR SURFACE BASED ON A COMBINED DEEP LEARNING AND SHAPE-FROM-SHADING (SFS) APPROACH

Hao Chen^{1,*}, Xuanyu Hu¹, Jürgen Oberst^{1,2}

¹ Institute of Geodesy and Geoinformation Science, Technische Universität Berlin, Kaiserin-Augusta-Allee 104-106, 10553 Berlin, Germany - hao.chen.2@campus.tu-berlin.de, (x.hu, juergen.oberst)@tu-berlin.de

² Institute of Planetary Research, German Aerospace Center (DLR), Rutherfordstr. 2, 12489 Berlin, Germany

Commission III, WG III/II

KEY WORDS: Pixel-resolution DTM, Lunar Surface, Deep Learning, Shape from Shading, Convolution Neural Network.

ABSTRACT:

High-resolution Digital Terrain Models (DTMs) of the lunar surface can provide crucial spatial information for lunar exploration missions. In this paper, we propose a method to generate high-quality DTMs based on a synthesis of deep learning and Shape from Shading (SFS) with a Lunar Reconnaissance Orbiter Narrow Angle Camera (LROC NAC) image as well as a coarse-resolution DTM as input. Specifically, we use a Convolutional Neural Network (CNN)-based deep learning architecture to predict initial pixel-resolution DTMs. Then, we use SFS to improve the details of DTMs. The CNN-model is trained based on the dataset with 30,000 samples, which are formed by stereo-photogrammetry derived DTMs and orthoimages using LROC NAC images as well as the Selenological and Engineering Explorer and LRO Elevation Model (SLDEM). We take Chang'E-3 landing site as an example, and use a 1.6 m resolution LROC NAC image and 5 m resolution stereo-photogrammetry derived DTM as input to test the proposed method. We evaluate our DTMs with those from stereo-photogrammetry and deep learning. The result shows the proposed method can generate 1.6 m resolution high-quality DTMs, which can clearly improve the visibility of details of the initial DTM generated from the deep learning method.

1. INTRODUCTION

Digital Terrain Models (DTMs) derived from high-resolution orbiter imagery represent fundamental spatial information to support lunar exploration missions. The resolution of images taken by Lunar Reconnaissance Orbiter Narrow Angle Camera (LROC NAC) images on board the Lunar Reconnaissance Orbiter (LRO) falls within 0.5 m – 2 m, which are the highest resolution orbiter images of the lunar surface available (Henriksen et al., 2017). Currently, stereo-photogrammetry, Shape from Shading (SFS), and Convolutional Neural Network (CNN)-based deep learning methods have been used to produce high-resolution DTMs based on the LROC NAC images. Deep learning and SFS techniques can use a single image and an initial existing low-resolution DTM to retrieve pixel-resolution DTMs (Grumpe et al., 2014; Liu et al., 2021). The performance of SFS is highly dependent on the quality of the input DTM (Wu et al., 2018). Currently, DTMs derived from laser altimetry or stereo-photogrammetry are the main source of input DTMs for SFS (Liu et al., 2020). Laser altimetry produces global DTMs with a high vertical accuracy, but the horizontal resolution is relatively coarse, for example, the Lunar Orbiter Laser Altimeter (LOLA) DEM with 30 m / pixel (Smith et al., 2010). Stereo-photogrammetry can generate 2 m – 5 m resolution DTMs based on LROC NAC stereo pairs (Henriksen et al., 2017). Although the LROC NAC images can almost cover the lunar surface completely, suitable stereo images that meet the requirements of conjugate matching and triangulation are spatially limited. It is therefore unlikely that stereo-photogrammetry can be used to produce high-resolution DTMs for all areas of interest, let alone as input to SFS.

CNN-based single depth estimation was introduced to estimate depth maps from single images in computer vision, and the performance has been steadily improved (Eigen et al., 2014; Alhashim et al., 2018). It has been applied to generate high- and even pixel-resolution DTMs using single images for the Martian surface (Chen et al., 2021; Tao et al., 2021a; Tao et al., 2021b; Tao et al., 2021c). Moreover, single depth completion technique was proposed, which can boost the predict accuracy significantly (Ma et al., 2018; Cheng et al., 2019). Specifically, in addition to images, sparse depth maps are also used as input to estimate high-resolution depth maps (Shivakumar et al., 2019). On the other hand, the CNN-based method tends to produce overly smoothed structures because of the large receptive field involved in the regularization (Yao et al., 2018; Chen et al., 2017; Chen et al., 2021). Despite this, the method is still able to retrieve finer terrain details than laser altimetry or stereo-photogrammetry. In addition, because deep learning requires only single images, it is not subject to the limitations of stereo-photogrammetry. Therefore, it is well suited as the initial DTM for the SFS method.

In this paper, we propose a DTM generation approach that combines deep learning and SFS. We present a data experiment to show that the method possesses not only the high elevation accuracy but also fine local structure details, which is superior to stereo-photogrammetry and deep learning methods.

2. METHOD

The principle of the proposed DTM generation approach is to use CNN-based deep learning derived DTMs as the initial

* Corresponding author

DTMs for SFS. The flow chart of the proposed work is shown in Figure 1. In this paper, we use CNN-based single depth completion technique as the deep learning method. First, during the pre-processing stage, we clip the size of the input according to the requirements of the method, and normalize these clipped data to ensure convergence of the CNN architecture. Then, the pre-processed data are used to train our CNN-based deep learning architecture, and we use the well-trained CNN model to predict the DTMs from single LROC NAC images and coarse-resolution DTMs. Subsequently, the predicted, normalized DTMs are recovered to the real scale. Finally, we use the LROC NAC images and the scale recovered DTMs as the input to SFS to refine the deep learning derived DTMs.

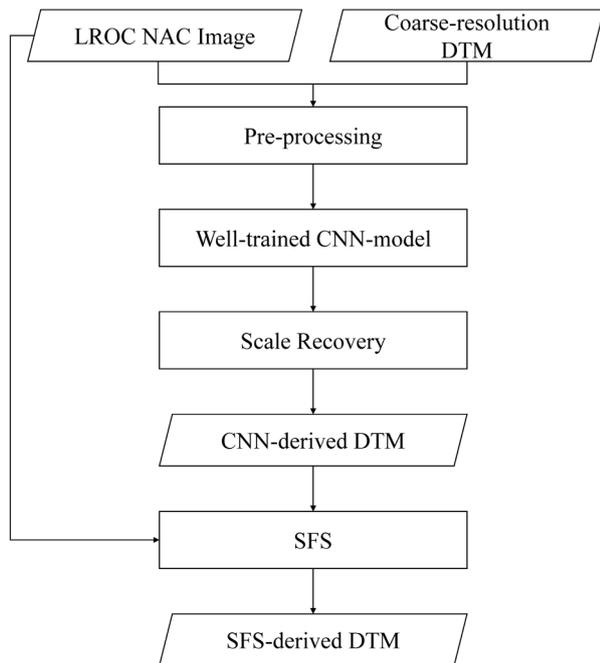


Figure 1. Flowchart of the DTM generation process.

2.1 Dataset and Pre-processing

The training and validation datasets for the deep learning method are formed from the public DTMs and Digital Orthophoto Maps (DOMs) produced via stereo-photogrammetry based on LROC NAC stereo pairs (downloaded from https://wms.lroc.asu.edu/lroc/rdr_product_select), and the corresponding Selenological and Engineering Explorer and LRO Elevation Model (SLDEM) with 512 pixels / degrees (downloaded from <http://imbrium.mit.edu/EXTRAS/SLDEM2015/>). As the SELENE DTM covers latitudes between 60° S/N, we only used NAC DTMs and orthoimage pairs from this region. Also, since their resolution falls into 2 m - 5 m, we individually cropped out the SLDEM for each NAC DTM and orthoimage pairs and interpolate them to the same resolution of the DTMs and DOMs. Considering the memory constraint of the graphic card used in our computing environment (see below), these data are clipped into 256 * 320 sub-images. We used the clipped LROC NAC DOMs and SLDEM as the respective input images and DTMs, the clipped LROC NAC DTMs as the ground truth.

The gray values of the images were regularized to [0, 1], and the elevations of DTMs were normalized to zero mean and unit standard deviation. Then, we subtracted the minimum value for

each normalized DTMs. We randomly selected 27, 000 samples as our training set, and 3,000 samples as our validation set at the end of the training. We carried out vertical and horizontal flips with 0.5 probability respectively during the training stage to expand the training set with more different illumination azimuth angle conditions.

2.2 Deep Learning Method

2.2.1 Network Architecture

In this study, we use a dual-encoder CNN architecture to estimate the pixel resolution DTM for the lunar surface. The network architecture is shown in Figure 2. It adopts two independent encoder branches to extract features from LROC NAC images and SLDEM, respectively, and uses a fuse module to concatenate the outputs from the two encoder modules into one volume. As the LROC NAC images contain fine texture information, the LROC NAC encoder module is based on the relatively complex ResNet50 architecture (He et al., 2016; Laina et al., 2016). In contrast, the SLDEM encoder module is based on five simple stacked convolution blocks to extract the elevation information. The decoder module is built with four up-projection blocks and a nearest upsampling block to recover the spatial resolution to match that of the input (Laina et al., 2016; Alhashim et al., 2018). The connections are added between the layers with the same spatial resolution in the LROC NAC encoder module and decoder module to increase the ability to predict more surface structure feature (Alhashim et al., 2018). Finally, a convolution layer is added to reduce the channels of the output from the nearest upsampling block to 1. This is the output of the model.

2.2.2 Loss Function

The weighted sum of the mean absolute error (L_1 ; Hambarde et al., 2020) and the gradient loss (L_{grad} ; Alhashim et al., 2018) are adopted to train the network. The L_1 loss can minimize the residual error between the predicted DTMs and the ground truth. The L_{grad} can account for the relationships between surrounding pixels to predict more correctly local surface features. The loss function (L) is defined by:

$$L = \lambda * L_1 + \gamma * L_{grad} \quad (1)$$

where λ and γ are the weight factors to be specified for the training

2.2.3 Scale Recovery

We initially recover the scale of the predicted DTM based on the heights of the reference coarse-resolution DTM, namely, the input DTM of the model:

$$DTM_{init} = v + \mu * DTM_{pred} \quad (2)$$

where DTM_{pred} = the predicted DTM from our model
 DTM_{init} = the initial scale recovered DTM
 v = the mean of the coarse-resolution DTM heights
 μ = the standard deviation of the coarse-resolution DTM heights

Then, we use the *pc_align* tool in Ames Stereo Pipeline (ASP) software to co-register the initial recovered DTMs to the reference DTM (Beyer et al., 2018).

2.2.4 Training of the Network Architecture

The network was implemented on PyTorch and trained using a single Nvidia® RTX3060® graphic card (Paszke et al., 2017). We used Adam optimizer for optimization (Kingma et al., 2014). The batch size was 4; the training epochs was 20. the initial learning rate was 0.0001; we successively reduced it to

10% after every 5 epochs; the loss function weights λ and γ were set to 1 and 0.5, respectively.

After training, we used the best model i.e., of the highest accuracy in reference to the validation set to predict DTMs. The outcomes were subsequently used as the initial DTMs for the following SFS method.

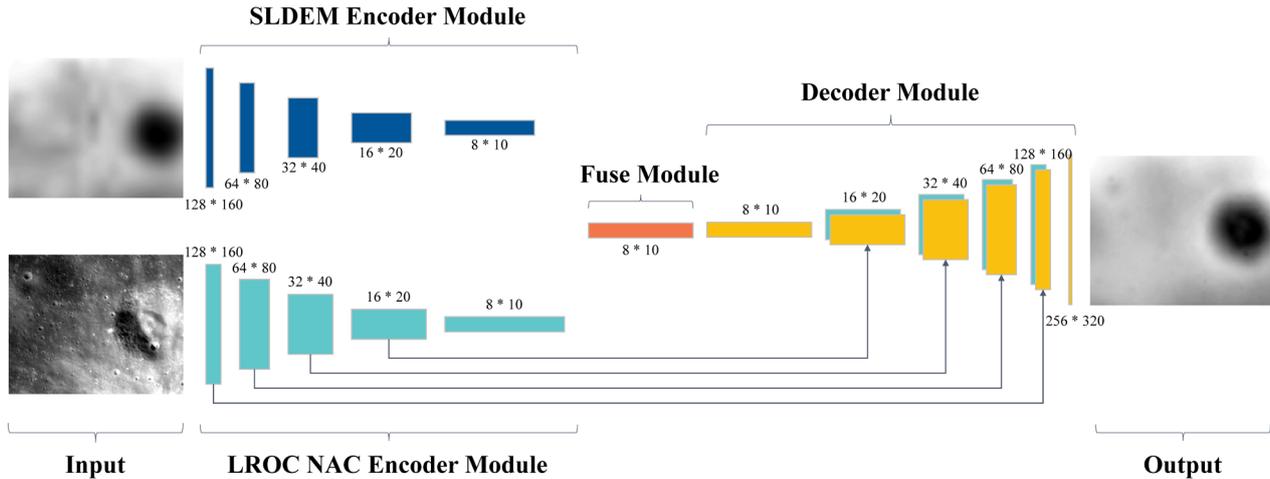


Figure 2. Overview of the network architecture. Each coloured box represents a block in the model. Numbers next to blocks in the Encoder module represent the input dimensions of the block. Numbers next to blocks in the Decoder module represent the output dimension.

2.3 SFS method

We use the LROC NAC images and the predicted DTMs via deep learning directly as inputs for SFS methods. In this paper, we simply use the *SFS* tool provided by the ASP to improve the details of the input DTMs. As we only consider a single image and an initial DTM as input, the code works by minimizing the following cost function (Alexandrov et al., 2018):

$$\iint \sum [P(Z)(x, y) - E\alpha(x, y)R(Z)(x, y)]^2 + \theta \|\nabla^2 Z(x, y)\|^2 + \kappa [Z(x, y) - Z_0(x, y)]^2 dx dy \quad (3)$$

where $Z(x, y)$ = the estimated DTM
 $P(Z)(x, y)$ = the camera image interpolated at pixels obtained by projecting into the camera 3D pointing from $Z(x, y)$
 E = the image exposure
 $\alpha(x, y)$ = the terrain-dependent albedo
 $R(Z)(x, y)$ = the reflectance computed from $Z(x, y)$ for the image
 θ = the smoothing weight
 $\|\nabla^2 Z(x, y)\|^2$ = the sum of squares of all second-order partial derivatives of $Z(x, y)$
 κ = the initial DTM constraint weight
 $Z_0(x, y)$ = the initial DTM

We selected the Lunar Lambertian model as the reflectance model of the local surface (McEwen et al., 1991). We set the initial smoothing weight as 0.08, and the initial DTM constraint weight as 0.001 in the following experiments.

3. EXPERIMENTS AND RESULTS

3.1 Test Data and Pre-processing

We used high-resolution LROC NAC images as the input image to test our approach for the Chang'E-3 landing site. We downloaded the LROC NAC EDR level images (M1144929211LE) from the PDS website. The illumination azimuth angle is 234.04°, and the illumination elevation angle is 33.11°. We used the Integrated System for Imagers and Spectrometers (ISIS) to process the EDR image to a map projected image with 1.6 m resolution. Also, we downloaded and used the 5 m resolution stereo-photogrammetry derived DTM as the input DTM from https://wms.lroc.asu.edu/lroc/view_rdr/NAC_DTM_CHANGE3. It was interpolated to the same resolution of the map projected LROC NAC image to unify the resolution.

Then, we took the Chang'E-3 landing site as the centre, and clipped the projected image and interpolated DTM to an area of 256 * 320 pixels in size (Figure 3). Then, the clipped image and DTM were regularized to [0, 1] and reduced to standard normal distribution as the input for the method, respectively.

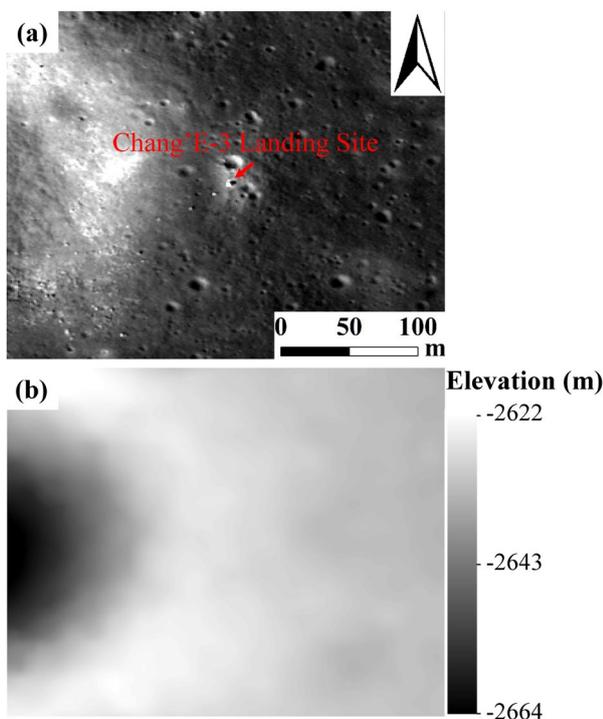


Figure 3. The test data in the Chang'E-3 landing site, (a) the input image, (a) the input DTM.

3.2 Results

Figure 4 shows the 5 m resolution stereo-photogrammetry derived DTM (Figure 4(d)), the 1.6 m resolution deep learning derived DTM (Figure 4(e)), the 1.6 m resolution SFS derived DTM (Figure 4(f)), the corresponding hill shaded maps and comparisons between elevation profiles across these three DTMs. Owing to the coarse resolution of stereo-photogrammetry DTM, its hill shaded map does not show small scale craters (Figure 4(a)). While the DTM can retrieve the

correct topographic trend, local areas exist apparently erroneous topographic undulations. The hill shaded map generated from deep learning DTM reveals small scale craters clearly. However, the recovered structures tend to be over-smoothed as described in Section 1. In contrast, the SFS derived DTM can show finer local details than the deep learning method, and these details are more consistent with the original LROC NAC images (Figure 4(c)). For example, several craters near the Chang'E-3 lander (cf. Fig. 3) can be accurately recovered (see the white arrows in Figure 4(c)), while the stereo-photogrammetry and deep learning method fail to retrieve them.

In the meantime, the profile comparison result shows the SFS-derived DTM is very well consistent with the other two DTMs (Figure 4(g)). Considering both terrain details and overall elevation accuracy, the SFS method based on the deep learning-derived initial DTM has the highest quality.

4. CONCLUSION

In this paper, we proposed a pixel-resolution DTM generation method for the lunar surface via deep learning and SFS. We used deep learning derived DTMs as input for SFS as a means of refinement. The result shows that the proposed approach can retrieve more detailed terrain structure information than those derived from deep learning and stereo-photogrammetry alone. Also, the overall elevation accuracy of the final SFS-refined DTM is inherently constrained by, and thus corresponds well to, the DTMs derived by deep learning or photogrammetry. Overall, the generated high-quality DTMs are suitable for high-precision landing site selection and assessment, or other applications which need high-resolution and precision spatial information.

As the next step, we will perform more tests for different areas and explore the fusion of multiple methods to generate higher accuracy DTMs.

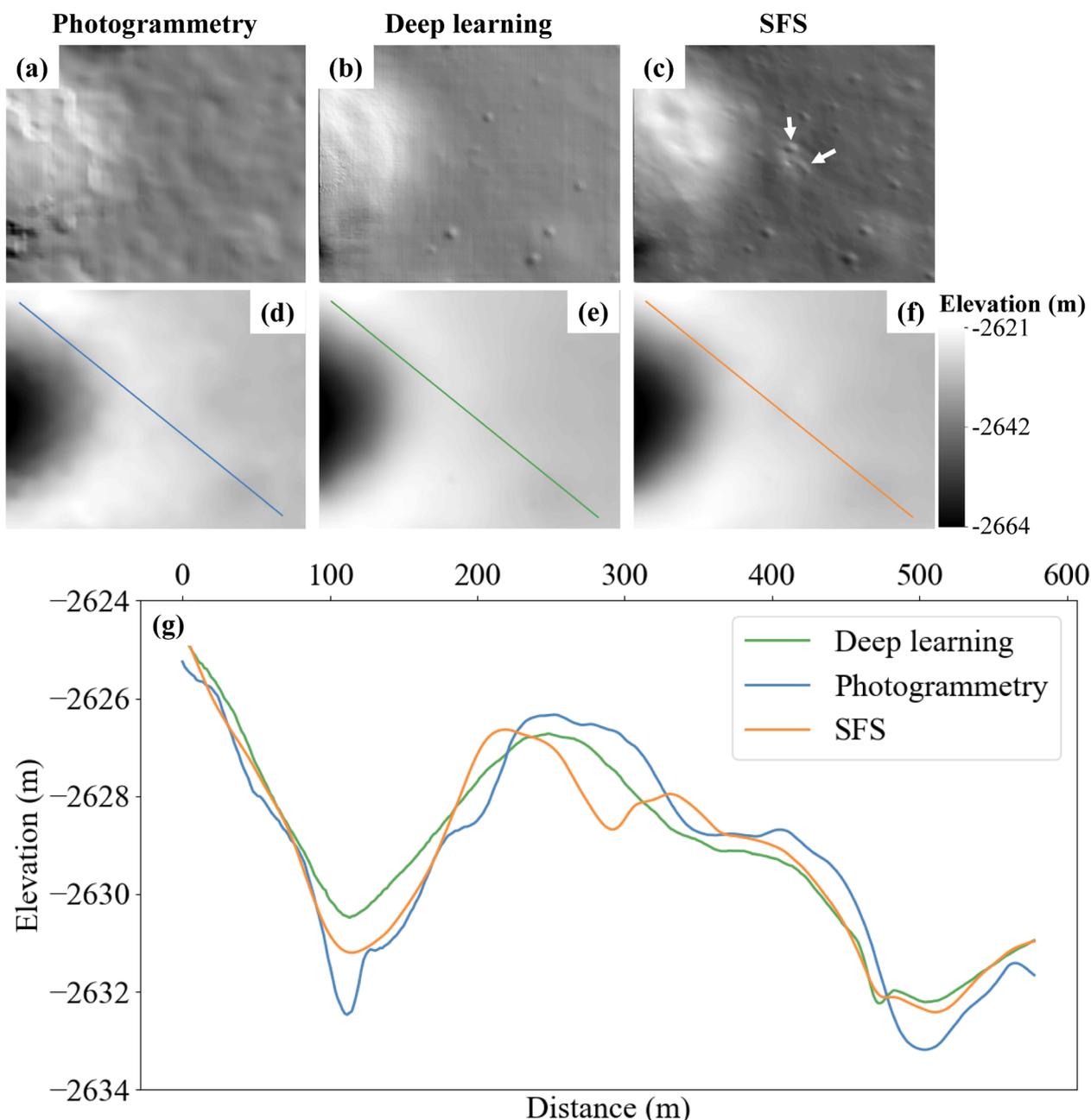


Figure 4. DTM generation results and comparison. The three top panels (a,b,c) show the hill shaded maps with the same illumination azimuth and elevation angles of the original LROC NAC images. The center panels (d,e,f) are the generated DTMs by stereo-photogrammetry, deep learning, and SFS, respectively. The bottom panel (g) shows the profile comparison between the three DTMs in middle panel. The location of the profile is shown on the DTMs. (For interpretation of the references in the legend, the reader is referred to the web (colour) version of this article.)

ACKNOWLEDGEMENTS

H. Chen thanks the China Scholarship Council (CSC) for the financial support of his Ph. D. study at Technische Universität Berlin, Germany. We thank all those who make the LROC NAC images, DOMs, and DTMs, as well as SLDEM publicly available.

REFERENCES

- Alhashim, I., Wonka, P., 2018. High quality monocular depth estimation via transfer learning. arXiv preprint arXiv:1812.11941.
- Alexandrov, O., Beyer, R. A., 2018. Multiview Shape- From-Shading for Planetary Images. *Earth and Space Science*, 5(10), 652-666.

- Barker, M. K., Mazarico, E., Neumann, G. A., Zuber, M. T., Haruyama, J., Smith, D. E., 2016. A new lunar digital elevation model from the Lunar Orbiter Laser Altimeter and SELENE Terrain Camera. *Icarus*, 273, 346-355.
- Beyer, R., Alexandrov, O., McMichael, S., 2018. NeoGeographyToolkit/StereoPipeline: NASA Ames Stereo Pipeline 2.6.1 (v2.6.1). Zenodo. <https://doi.org/10.5281/zenodo.1345235>.
- Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4), 834-848.
- Chen, Z., Wu, B., Liu, W. C., 2021. Mars3DNet: CNN-based high-resolution 3D reconstruction of the Martian surface from single images. *Remote Sensing*, 13(5), 839.
- Cheng, X., Wang, P., Yang, R. 2019. Learning depth with convolutional spatial propagation network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(10), 2361-2379.
- Eigen, D., Puhrsch, C., Fergus, R., 2014. Depth map prediction from a single image using a multi-scale deep network. *arXiv preprint arXiv:1406.2283*.
- Grumpe, A., Belkhir, F., Wöhler, C., 2014. Construction of lunar DEMs based on reflectance modelling. *Advances in Space Research*, 53(12), 1735-1767.
- Hambarde, P., Murala, S., 2020. S2dnet: Depth estimation from single image and sparse samples. *IEEE Transactions on Computational Imaging*, 6, 806-817.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778.
- Henriksen, M. R., Manheim, M. R., Burns, K. N., Seymour, P., Speyerer, E. J., Deran, A., Boyd, A. K., Howington-Kraus, E., Rosiek, M. R., Archinal, B. A., Robinson, M. S., 2017. Extracting accurate and precise topography from LROC Narrow Angle Camera stereo observations. *Icarus*, 283, 122-137.
- Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Laina, I., Rupprecht, C., Belagiannis, V., Tombari, F., Navab, N., 2016. Deeper depth prediction with fully convolutional residual networks. In *2016 Fourth International Conference on 3D Vision*, 239-248.
- Liu, W. C., Wu, B., 2020. An integrated photogrammetric and photoclinometric approach for illumination-invariant pixel-resolution 3D mapping of the lunar surface. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159, 153-168.
- Liu, W. C., Wu, B., 2021. Influence of solar incidence angle on single-image photoclinometry for precision lunar topographic mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 182, 208-227.
- Ma, F., Karaman, S., 2018. Sparse-to-dense: Depth prediction from sparse depth samples and a single image. In *2018 IEEE International Conference on Robotics and Automation*, 4796-4803.
- McEwen, A. S., 1991. Photometric functions for photoclinometry and other applications. *Icarus*, 92(2), 298-311.
- Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A., 2017. Automatic differentiation in PyTorch.
- Shivakumar, S. S., Nguyen, T., Miller, I. D., Chen, S. W., Kumar, V., Taylor, C. J., 2019. Dfusenet: Deep fusion of rgb and sparse depth information for image guided dense depth completion. In *2019 IEEE Intelligent Transportation Systems Conference*, 13-20.
- Smith, D. E., Zuber, M. T., Jackson, G. B., Cavanaugh, J. F., Neumann, G. A., Riris, H., Sun, X., Zellar, R. S., Coltharp, C., Connelly, J., Katz, R. B., Kleyner, I., Liiva, P., Matuszeski, A., Mazarico, E. M., McGarry, J. F., Novo-Gradac, A., Ott, M. N., Peters, C., Ramos-Izquierdo, L. A., Ramsey, L., Rowlands, D. D., Schmidt, S., Stanley Scott III, V., Shaw, G. B., Smith, J. C., Swinski, J., Torrence, M. H., Unger, G., Yu, A. W., Zagwodzki, T. W., 2010. The Lunar Orbiter Laser Altimeter investigation on the Lunar Reconnaissance Orbiter mission. *Space Science Reviews*, 150(1), 209-241.
- Tao, Y., Xiong, S., Conway, S. J., Muller, J. P., Guimpier, A., Fawdon, P., Thomas, N., Cremonese, G., 2021a. Rapid single image-based DTM estimation from ExoMars TGO CaSSIS images using Generative Adversarial U-Nets. *Remote Sensing*, 13(15), 2877.
- Tao, Y., Muller, J. P., Conway, S. J., Xiong, S., 2021b. Large area high-resolution 3D mapping of Oxia Planum: The landing site for the ExoMars Rosalind Franklin rover. *Remote Sensing*, 13(16), 3270.
- Tao, Y., Muller, J. P., Xiong, S., Conway, S. J., 2021c. MADNet 2.0: pixel-scale topography retrieval from single-view orbital imagery of Mars using deep learning. *Remote Sensing*, 13(21), 4220.
- Wu, B., Liu, W. C., Grumpe, A., Wöhler, C., 2018. Construction of pixel-level resolution DEMs from monocular images by shape and albedo from shading constrained with low-resolution DEM. *ISPRS Journal of Photogrammetry and Remote Sensing*, 140, 3-19.
- Yao, Y., Luo, Z., Li, S., Fang, T., Quan, L., 2018. Mvsnet: Depth inference for unstructured multi-view stereo. In *Proceedings of the European Conference on Computer Vision*, 767-783.